A New Statistical Approach for Ozone Prediction with Quantification of Forecast Uncertainty

PENNSTATE 1855

Nikolay V. Balashov[1], Anne M. Thompson[1,2], George S. Young[1]

[1]Penn State University – College of Earth and Mineral Science – Department of Meteorology

[2]NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771

(Author Email: nvb5011@psu.edu)

NASA

INTRODUCTION

- The use of statistical methods to help predict air quality has been shown to be an effective tool for air quality (AQ) forecasters (Garner and Thompson, 2013; EPA 2013; Zhang et al., 2012)
- Statistical methods allow AQ forecasters to identify the magnitude of uncertainty associated with a forecast enabling them to make a better AQ prediction (Garner and Thompson, 2013)

GOALS

- Develop a novel stand-alone (without help of chemistry transport models) statistical approach to predict 8-h ozone maximum that specifies forecast uncertainty with the help of a clustering algorithm known as self-organizing maps (SOMs)

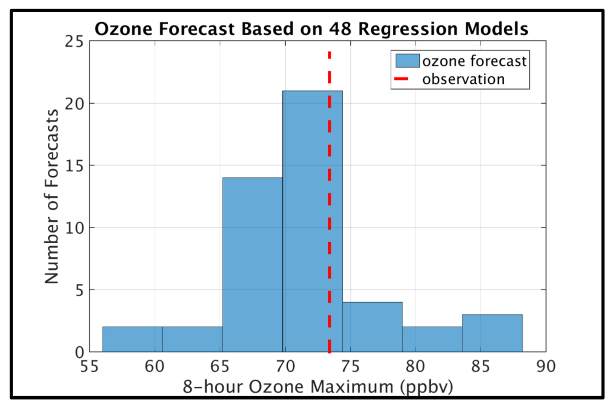


Figure 1 (right). *An example of a product that quantifies uncertainty in an 8-hour ozone maximum forecast.*

- Evaluate the approach using 2013 June-July-August (JJA) ozone data from a station in the San Joaquin Valley, CA – a US region that sees the most ozone exceedances

BACKGROUND

- Statistical models work by quantifying relationships between predictors such as meteorological variables and 8-h ozone maximum values (EPA, 2003)
- We develop our model using an ozone dataset from a representative monitoring site in the San Joaquin Valley
- To train our model we use 1987-2012 June-July-August (JJA) detrended ozone data (to compensate for significant NO_x emission changes over this period) measured at the Parlier AQ station
- Local meteorological variables are from the Fresno station which is a weather proxy for Parlier

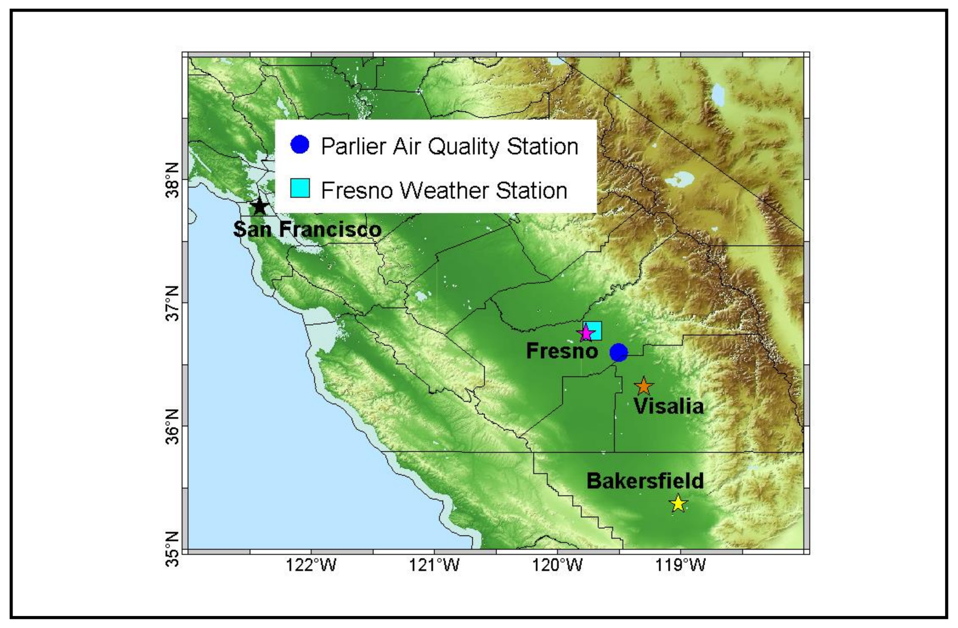


Figure 2 (above). *Map of the San Joaquin Valley showing major regional cities along with the air quality and weather stations that are used in this work.*

OZONE FORECASTING MODEL

- The model uses 3 different but related methods to make 3 distinct single-value 8-h ozone maximum forecasts
- These 3 forecasts can be averaged to create one consensus forecast
- Two of the methods are able to generate probability density functions (PDFs) of the forecast, providing information on forecast uncertainty

Method 1: Regression (REG)

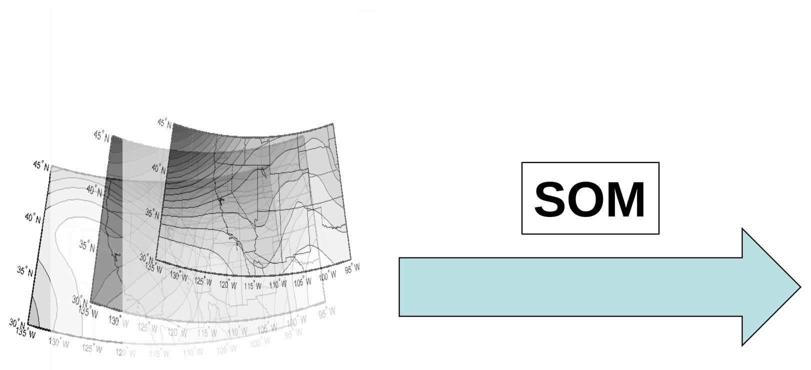
- Quadratic Stepwise Regression:

$$y_{o3} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \cdots + \epsilon$$

- Predictors (x) are temperature, wind speed, wind direction and previous day ozone

Method 2: Regression in SOM (REGiS)

- *Synoptic setting* modulates ozone response to local weather and chemical variables
- Therefore categorizing the *synoptic setting* decreases modeled-ozone variance and allows for more accurate statistical models based on local predictors
- SOM, which is a powerful clustering technique, is used to identify a representative number of *synoptic settings* (distinct weather regimes) for an AQ forecast region (Figure 3)
- Once the weather regimes are established, REGiS generates a regression equation (similar to the one in method 1) for each *synoptic setting* (Figures 4a and 4b)



Daily 500 hPa geopotential height, 2m T, 10m V, 10m U, and 850 hPa T fields from 1987-2012

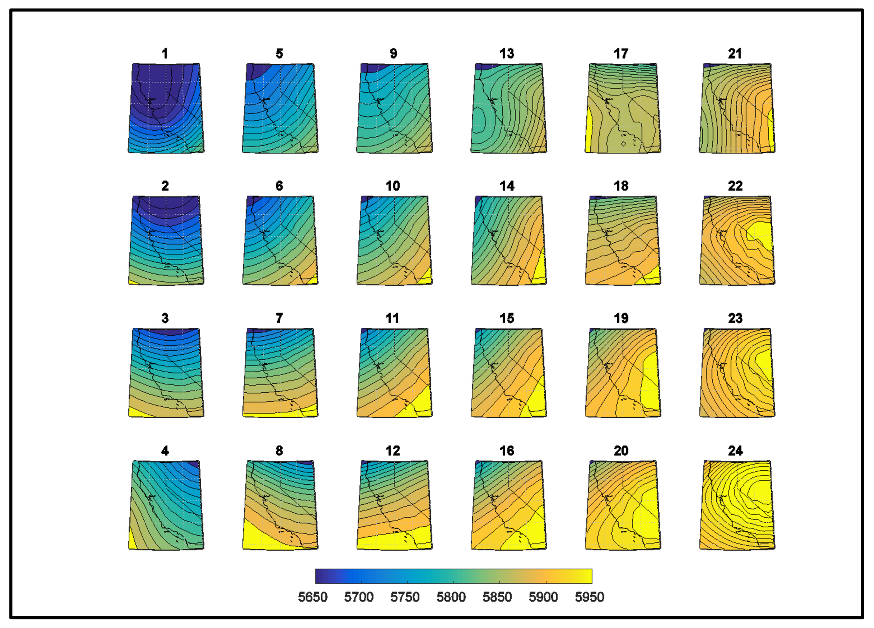
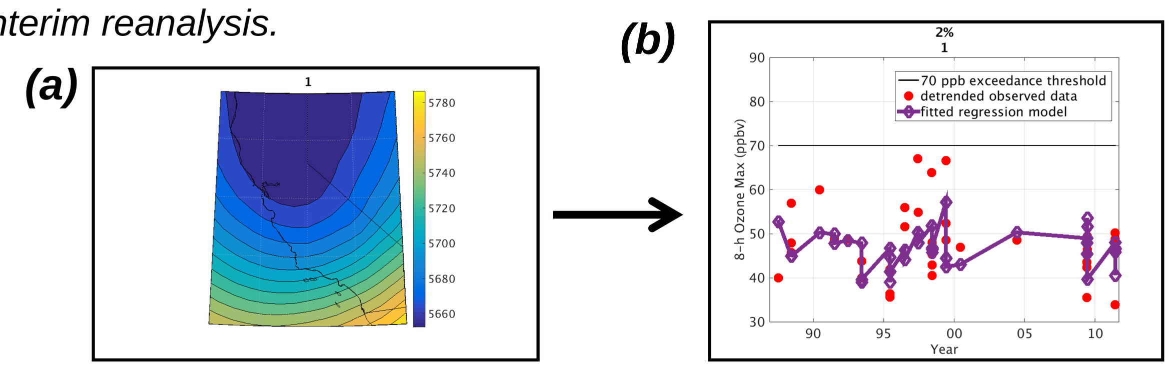


Figure 3 (above right). *4x6 SOM of daily 500 hPa geopotential height, 2m T, 10m V, 10m U, and 850 hPa T fields from June-July-August over 1987-2012, centered on the Western United States. Note: only 500 hPa geopotential height fields (in meters) are shown in full SOM. The data used for this analysis comes from ERA-Interim reanalysis.*

Figure 4 (right). ***(a)*** *A weather regime identified by SOM.* ***(b)*** *Regression model for a specific weather regime.*



Predict weather regime by a numerical weather model → Match the predicted regime with categorized regimes → Weight forecasts based on similarity between predicted and categorized regimes → Generate a PDF of the forecast

Figure 5 (above). *REGiS prediction procedure schematic.*

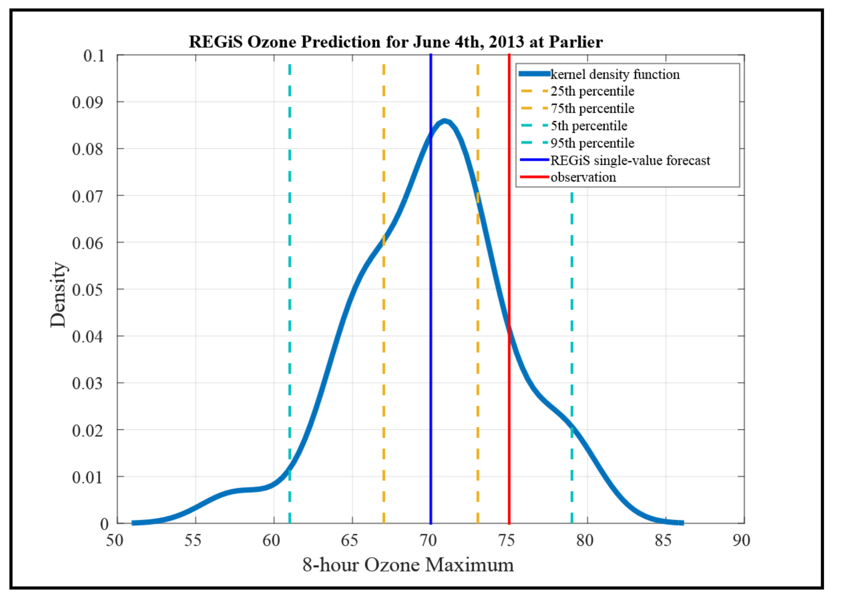
Method 3: SOM in SOM (SiS)

- Use SOM to identify distinct weather regimes for a region of an AQ forecast (as shown in **Figure 3**)
- For each weather regime, cluster predictors into distinct groups. Basically use SOM again but now on predictors (predictors as in method 1) within each regime

RESULTS

- Both REGiS and SiS are able to produce PDFs of an 8-hour maximum ozone forecast (Figures 6 and 7)
- The PDFs can be supplemented with single-value forecasts from REG, REGiS, SiS, and their consensus (Figure 7)

Figure 6 (right). *PDF of 8-h ozone maximum forecast generated by REGiS at Parlier for June 4st, 2013. Dashed lines indicate 25-75 percentile (orange) and 5-95 (blue) percentile regions. REGiS single-value forecast is in blue and observation is in red.*



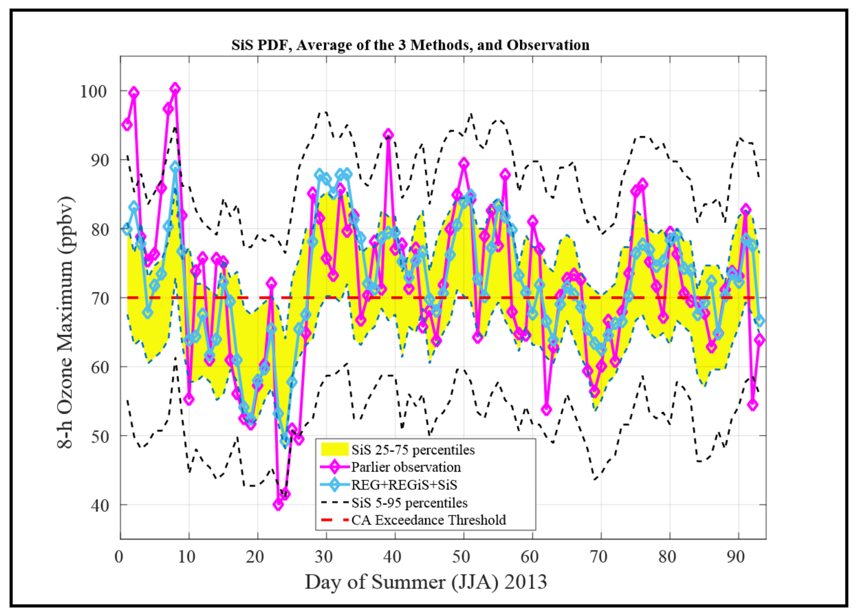


Figure 7 (left). *SiS PDF and consensus forecast vs. observations at Parlier for JJA 2013.*

Table 1. *RMSE for each method.*

REG	**8.16**
REGiS	**7.91**
SiS	**9.10**
Consensus	**7.68**

CONCLUSIONS

- Two separate but related ozone forecasting methods, REGiS and SiS, that are based on *synoptic setting* identification are developed using SOM clustering algorithm
- Clustering based on *synoptic setting* reduces the modeled-ozone variance making the regression model more effective at predicting the ozone
- REGiS and SiS are combined with quadratic stepwise regression (REG) to yield 3 distinct single-value ozone forecasts that contain some independent information with regard to the foreacst
- Consensus forecast exploits this independent information to yield forecast skill greater than that of each individual approach
- PDFs produced by REGiS and SiS allow for a quantification of ozone forecast uncertainty
- In our future work we will develop a similar model for PM2.5

ACKNOWLEDGMENTS

Gator research group. NASA for their enlightening DISCOVER-AQ campaign experience. Research was supported by NASA through Grants to Penn State University, DISCOVER-AQ (NNX10AR39G) and Applied Sciences Air Quality (NNX11AQ44G).

SELECTED REFERENCES

EPA, U.S., 2003. Guidelines for developing an air quality (Ozone and PM2.5) forecasting program, U.S. Environmental Protection Agency.

Garner, G.G. and Thompson, A.M., 2013. Ensemble statistical post-processing of the National Air Quality Forecast: Enhancing ozone forecasts in Baltimore, Maryland. Atmospheric Environment, 81(0): 517-522.

Kohonen, T., 2013. Essentials of the self-organizing map. Neural Networks, 37: 52-65.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C. and Baklanov, A., 2012. Real-time air quality forecasting, part I: History, techniques, and current status. Atmospheric Environment, 60(0): 632-655.